共変量判別問題における漸近展開による誤確率 (サンプリングの数理的研究)

山口, 光代

数理解析研究所講究録
272巻 1976年 81-88

共変量判別問題における漸近展開による誤確率

阪大. 基礎工学部　山口光代

q次元の共変量を持つ、2つの、$p+q$次元正規分布
$\Pi_1: N\left[\begin{bmatrix}\mu_1\\ \eta\end{bmatrix}, \Sigma\right]$、$\Pi_2: N\left[\begin{bmatrix}\mu_2\\ \eta\end{bmatrix}, \Sigma\right]$ のどちらか一方から、とりだされたことがわかっている観測値$\begin{bmatrix}x\\ y\end{bmatrix}$を、$\Pi_1, \Pi_2$からとり出された、大きさ$N_1, N_2$の2つの標本

$$\begin{bmatrix}x_{11}\\ y_{11}\end{bmatrix}, \cdots, \begin{bmatrix}x_{1N_1}\\ y_{1N_1}\end{bmatrix} \quad ; \quad \begin{bmatrix}x_{21}\\ y_{21}\end{bmatrix}, \cdots, \begin{bmatrix}x_{2N_2}\\ y_{2N_2}\end{bmatrix}$$

に基づいて、どちらかに判別する問題を考える。ここに、p次元平均ベクトルμ_1, μ_2，共通のq次元平均ベクトルη，共通のnonsingularな共分散行列Σは，すべて未知パラメーターである。この判別問題に対して、与えられている2つの判別統計量を用いる。1つは、Cochran and Bliss [1] による

$$W^* = (\bar{x}_1^* - \bar{x}_2^*)' S_{11\cdot 2}^{-1} [x^* - \tfrac{1}{2}(\bar{x}_1^* + \bar{x}_2^*)]$$

他の1つは、Fujikoshi and Kanazawa [2] による

$$Z^* = N_1 (N_1 + 1 + l_1/k_1)^{-1} (x^* - \bar{x}_1^*)' S_{11\cdot 2}^{-1} (x^* - \bar{x}_1^*)$$
$$- N_2 (N_2 + 1 + l_2/k_2)^{-1} (x^* - \bar{x}_2^*)' S_{11\cdot 2}^{-1} (x^* - \bar{x}_2^*)$$

である。$x^* = x - S_{12}S_{22}^{-1}y$, $\bar{x}_i^* = \bar{x}_i - S_{12}S_{22}^{-1}\bar{y}_i$, $\bar{x}_i$, $\bar{y}_i$：標本平均, $S_{11\cdot 2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$, $S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$ は Σ の最良不偏推定量, $l_i = (y-\bar{y}_i)'S_{22}^{-1}(y-\bar{y}_i)$, $k_i = n/N_i$, $n = N_1 + N_2 - 2$.

W^* による判別方式, Z^* による判別方式は、それぞれ,

$W^* \geq c_1 \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix}$ を π_1 に判別, $W^* < c_1 \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix}$ を π_2 に判別せよ.

$Z^* < c_2 \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix}$ を π_1 に判別, $Z^* \geq c_2 \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix}$ を π_2 に判別せよ.

となり, c_1, c_2 は、定数, 又は, $\bar{x}_i$, $\bar{y}_i$, S の関数である。この判別方式によって, $\begin{bmatrix} x \\ y \end{bmatrix}$ を判別するとき, 二種類の誤判別が起こるが, その中の一つの誤判別の確率, 即ち, $\begin{bmatrix} x \\ y \end{bmatrix}$ が実際は、π_1 に属するのに、誤って π_2 に属すると判別する確率

$$\Pr\{W^* < c_1 \mid \begin{bmatrix} x \\ y \end{bmatrix} \in \pi_1\}, \quad \Pr\{Z^* \geq c_2 \mid \begin{bmatrix} x \\ y \end{bmatrix} \in \pi_1\}$$

が一定の値, α, になるように、分離点 c_1, c_2 を定める場合, を考える。Studentized-W^*, Studentized-Z^* の分布関数の漸近展開を用いると, c_1, c_2 は, $(N_1, N_2, n)^{-2}$ 次の項を無視すると, 次の様に与えられる。$\Delta^2 = (\mu_1-\mu_2)'\Sigma_{11\cdot 2}^{-1}(\mu_1-\mu_2)$ として.

$$c_1 = c_\alpha(W^*) = A(W^*)D + \frac{1}{2}D^2, \quad c_2 = c_\alpha(Z^*) = 2A(Z^*)D - D^2,$$

$$A(W^*) = u_0 - (2N_1\Delta)^{-1}\{\Delta u_0 - 2(p-1)\} - (4n)^{-1}\{u_0^3 + (4p+4q-3)u_0\},$$

$$A(Z^*) = -u_0 - (2N_1\Delta)^{-1}\{u_0^2 + \Delta u_0 - (p-1)\} + (2N_2\Delta)^{-1}[u_0^2 + 2\Delta u_0 + \Delta^2 + p - 1]$$
$$- (4n)^{-1}\{u_0^3 + (4p-3)u_0 - 2q\Delta\}. \quad D^2 = (\bar{x}_1^* - \bar{x}_2^*)'S_{11\cdot 2}^{-1}(\bar{x}_1^* - \bar{x}_2^*).$$

これらの分離点に対して, もう一方の誤判別の確率, 即ち, $\begin{bmatrix} x \\ y \end{bmatrix}$ が実際は, π_2 に属するのに, 誤って π_1 に属すると判別する

確率は，それぞれ，

$$\Pr\{W^* \ge c_\alpha(W^*) \mid [\begin{smallmatrix}x\\y\end{smallmatrix}] \in \Pi_2\} = 1-\Phi(u_0+\Delta)+\phi(u_0+\Delta)\,[\,(2N_1\Delta)^{-1}(\Delta^2+p-1)$$
$$+(2N_2\Delta)^{-1}\{4\Delta(u_0+\Delta)-3(p-1)\}+(4n)^{-1}\Delta\{u_0^2+2(p+q-1)\} + O_2\,],$$

$$\Pr\{Z^* < c_\alpha(Z^*) \mid [\begin{smallmatrix}x\\y\end{smallmatrix}] \in \Pi_2\} = 1-\Phi(u_0+\Delta)+\phi(u_0+\Delta)\,[\,-(2N_1\Delta)^{-1}\{2u_0^2$$
$$+4\Delta u_0+3\Delta^2+p-1\}+(2N_2\Delta)^{-1}\{2(u_0+\Delta)(u_0+3\Delta)-(p-1)\}+(4n)^{-1}\Delta\{u_0^2+$$
$$2(p-q-1)\}+O_2\,]$$

$\Phi(u)$, $\phi(u)$ は $N[0,1]$ の cdf, pdf, u_0 は $\Phi(u)$ の 100α % 点，

となる．　小さい確率を与える統計量は，

Δ	$k_2 \le k_1$	$k_1 < k_2 < 2k_1+q$	$2k_1+q = k_2$	$2k_1+q < k_2$	
				$u_0 < u^{**}$	$u^{**} \le u_0$
0 ↓ 大	Z^*	W^*	W^*	W^*	W^*
		Δ_4		Δ_4	
		Z^*	Δ_0	Z^*	
			Z^*	Δ_3	
				W^*	

$$\Delta_0 = -(u_0^2+p-1)/2u_0$$

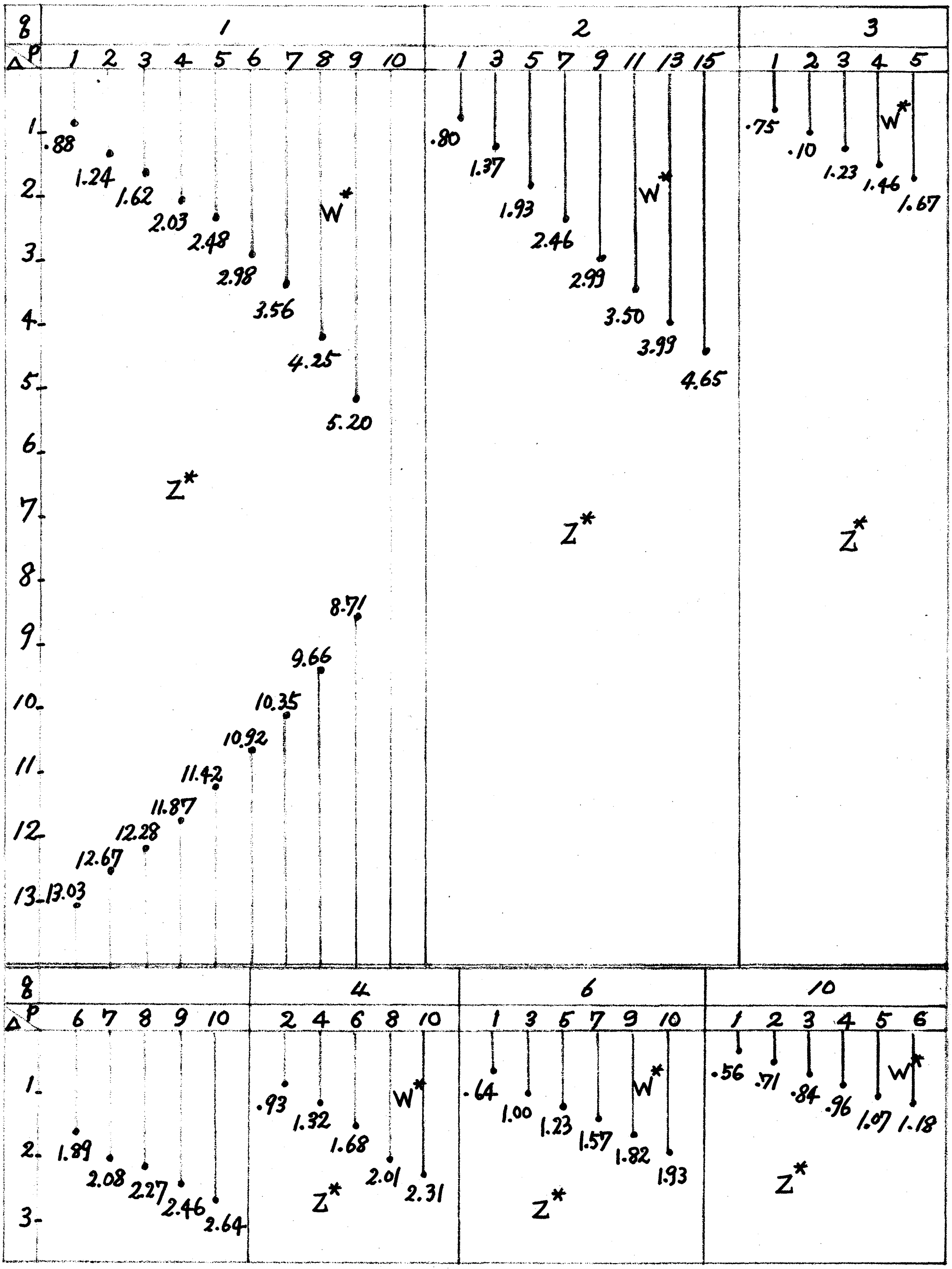
$$\Delta_3 = \{-(k_2-k_1)u_0+\sqrt{D_2}\}/(k_2-2k_1-q)$$

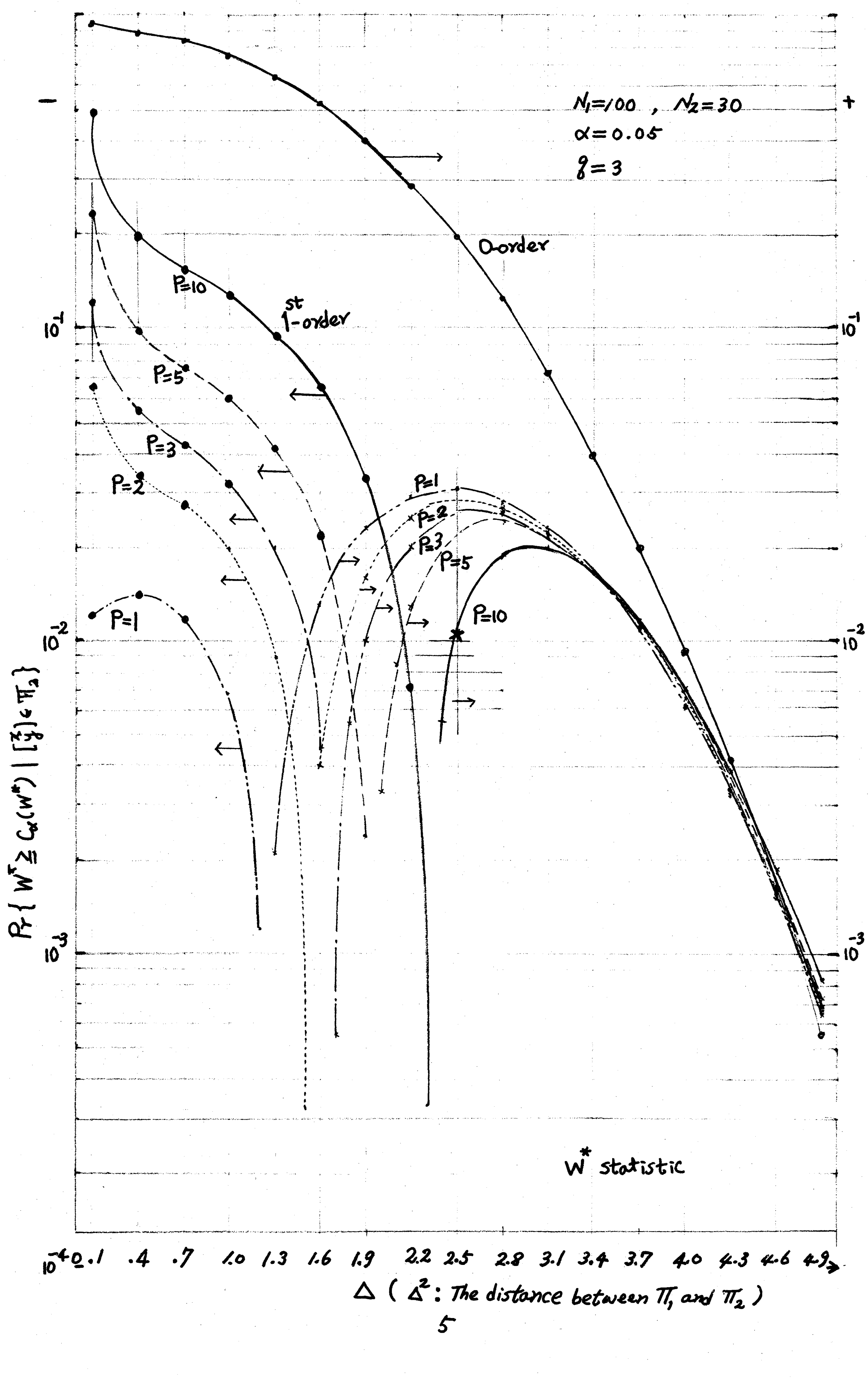
$$\Delta_4 = \{-(k_2-k_1)u_0-\sqrt{D_2}\}/(k_2-2k_1-q)$$

$$D_2 = (k_2-k_1)\{(k_1+q)u_0^2-(k_2-2k_1-q)(p-1)\}$$

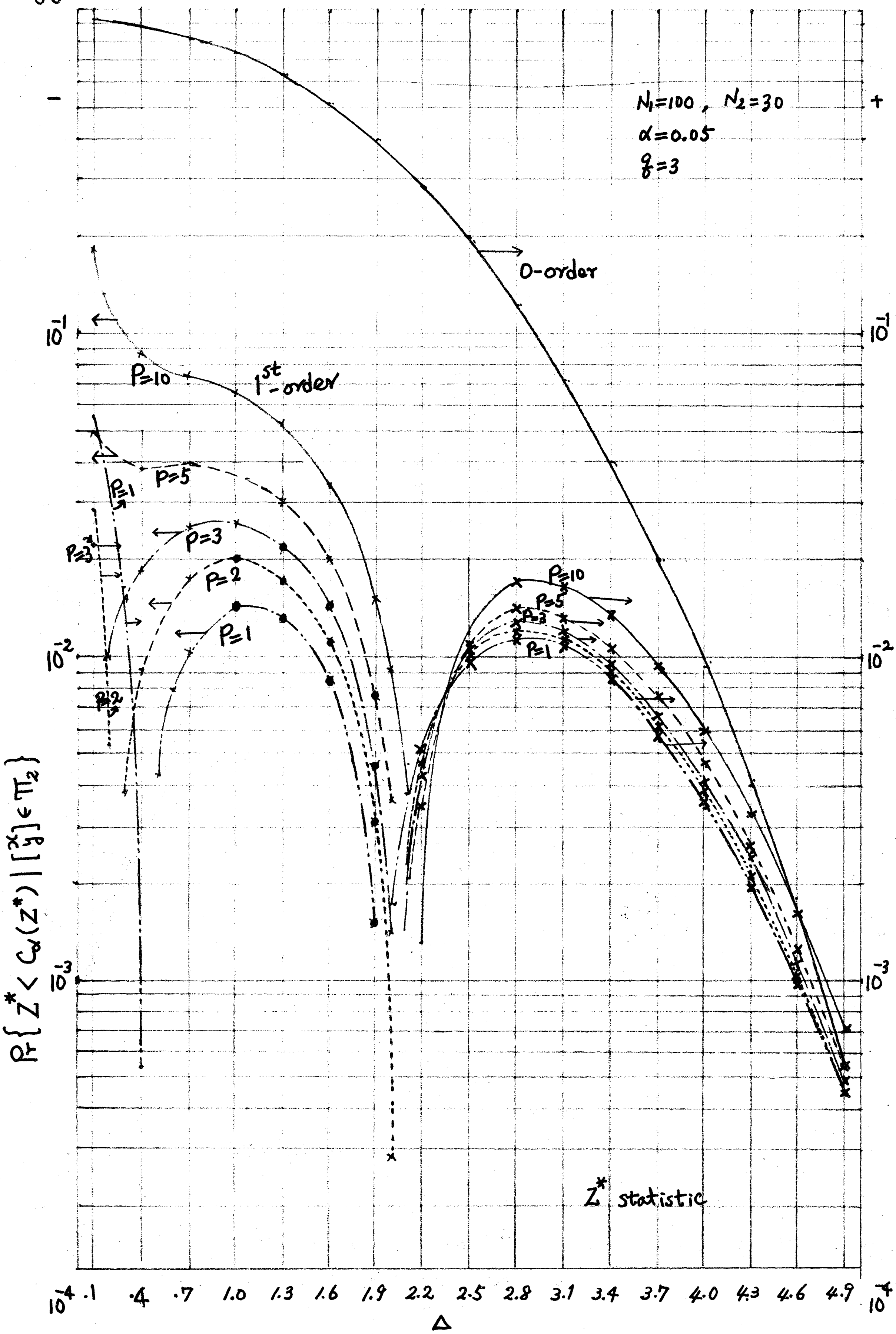
$$u^{**} = -\{(k_2-2k_1-q)(p-1)/(k_1+q)\}^{1/2}.$$

$N_1=100$, $N_2=30$, $\alpha=0.05$ に対して.

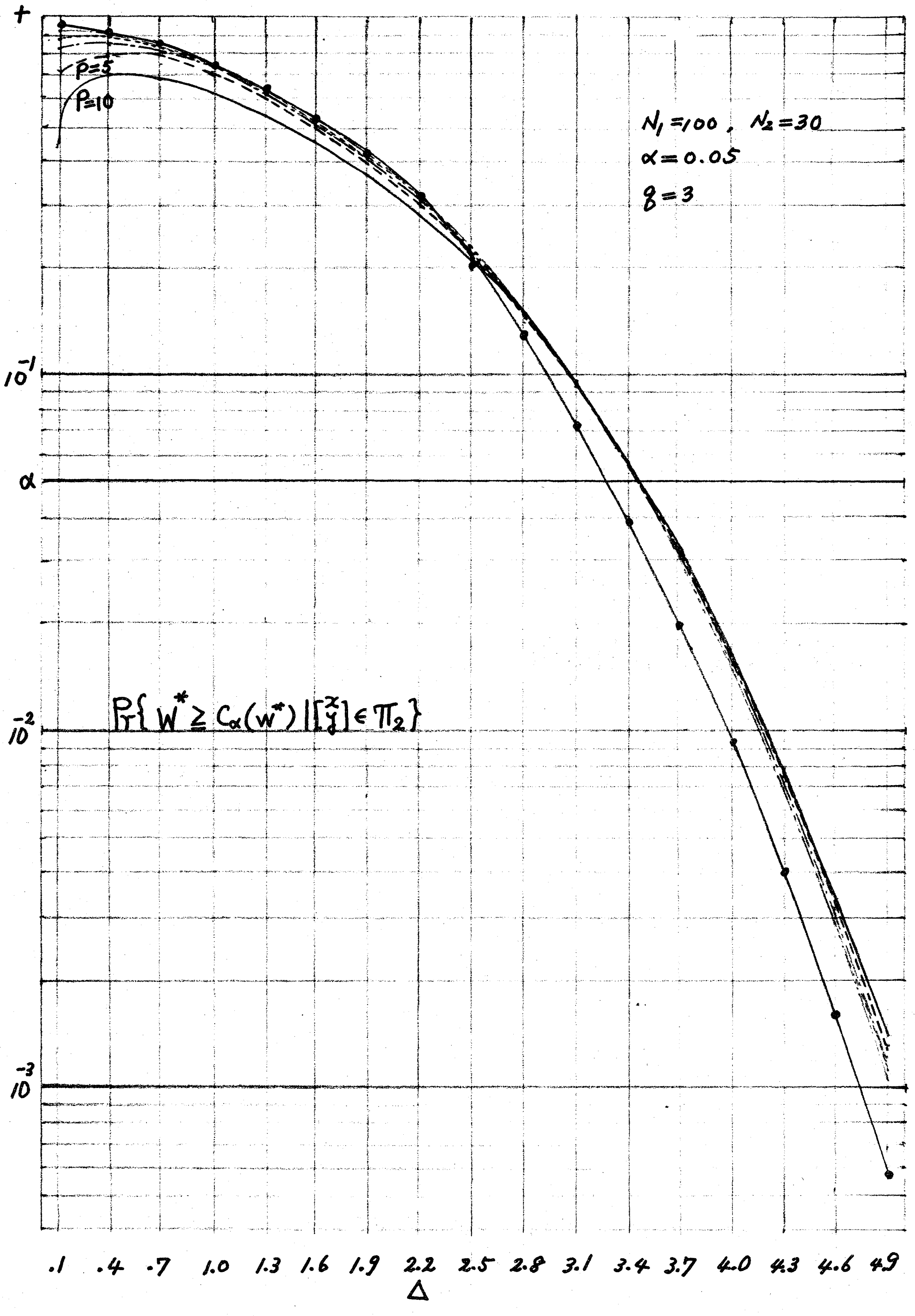



$N_1=100$, $N_2=30$
$\alpha=0.05$
$q=3$
0-order
1st-order
P=10
P=5
P=3
P=2
P=1
P=1
P=2
P=3
P=5
P=10
10^{-1}
10^{-2}
10^{-3}
10^{-4}
W^* statistic
$Pr\{W^* \geq C_\alpha(W^*) \mid [{x \atop y}] \in \Pi_2\}$
0.1 .4 .7 1.0 1.3 1.6 1.9 2.2 2.5 2.8 3.1 3.4 3.7 4.0 4.3 4.6 4.9
Δ (Δ^2: The distance between Π_1 and Π_2)

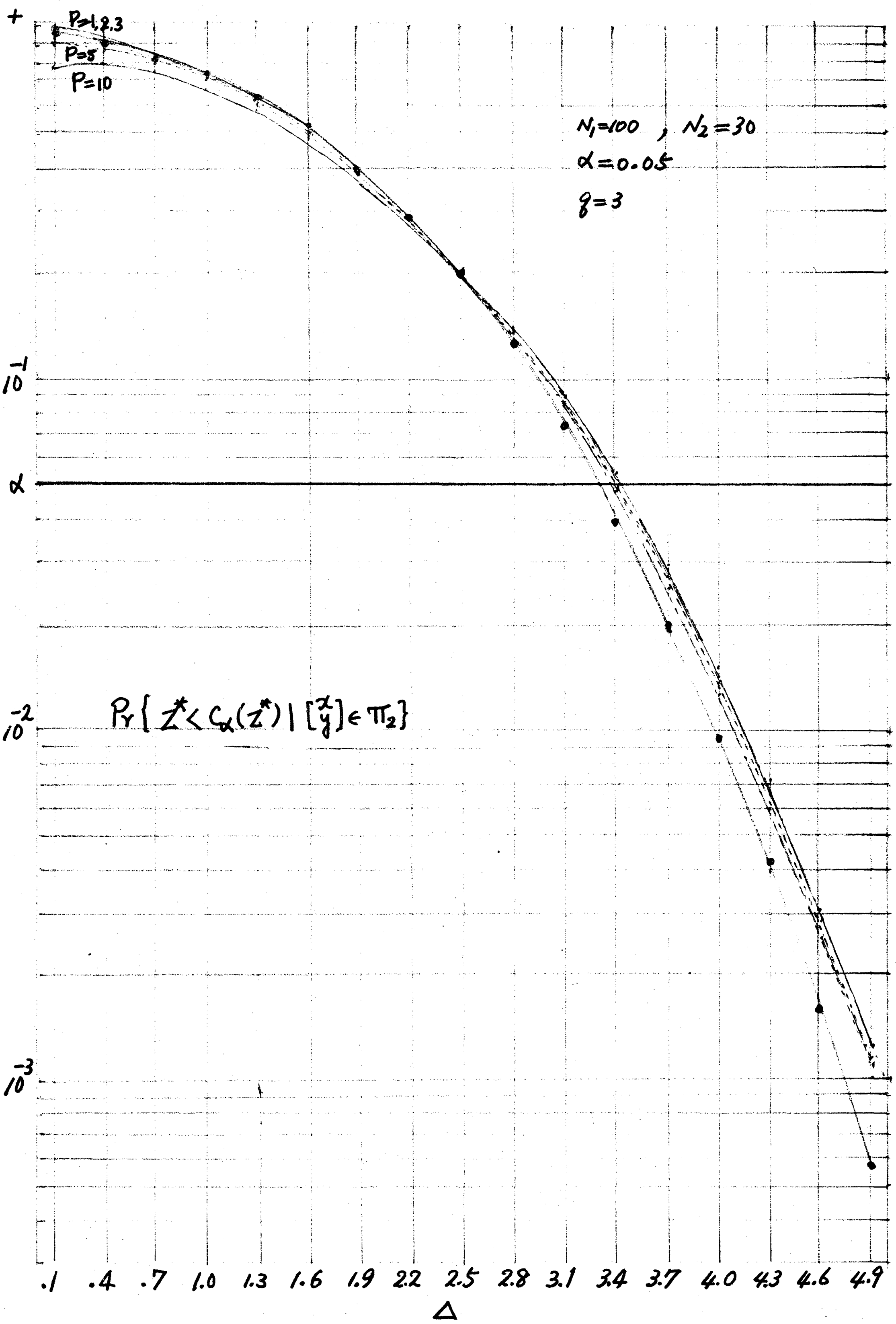
5


$N_1=100$, $N_2=30$
$\alpha=0.05$
$q=3$
0-order
P=10
1^{st}-order
P=5
P=1
P=3
P=2
P=1
P=10
P=5
P=3
P=1
Z^* statistic
$Pr\{Z^* < C_\alpha(Z^*) \mid [\begin{smallmatrix}x\\y\end{smallmatrix}] \in \pi_2\}$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
.1 .4 .7 1.0 1.3 1.6 1.9 2.2 2.5 2.8 3.1 3.4 3.7 4.0 4.3 4.6 4.9
Δ

6


+
P=5
P=10
$N_1 = 100$, $N_2 = 30$
$\alpha = 0.05$
$q = 3$
10^{-1}
α
10^{-2}
$Pr\{W^* \geq C_\alpha(W^*) \mid [\begin{smallmatrix} x \\ y \end{smallmatrix}] \in \Pi_2\}$
10^{-3}
.1 .4 .7 1.0 1.3 1.6 1.9 2.2 2.5 2.8 3.1 3.4 3.7 4.0 4.3 4.6 4.9
Δ

7



P=1,2,3
P=5
P=10
$N_1=100$, $N_2=30$
$\alpha=0.05$
$q=3$
10^{-1}
α
10^{-2}
10^{-3}
$Pr\{Z^* < C_\alpha(Z^*) \mid [\begin{smallmatrix}x\\y\end{smallmatrix}] \in \pi_2\}$
.1 .4 .7 1.0 1.3 1.6 1.9 2.2 2.5 2.8 3.1 3.4 3.7 4.0 4.3 4.6 4.9
Δ

8